DISEI - Università degli Studi di Firenze

Working Papers - Economics

Upward and downward bias when measuring inequality of opportunity

P. Brunori, V. Peragine, L. Serlenga

Working Paper N. 02/2017

DISEI, Università degli Studi di Firenze
Via delle Pandette 9, 50127 Firenze, Italia
www.disei.unifi.it

Upward and downward bias when measuring inequality of opportunity

Paolo Brunori[*], Vito Peragine[†], Laura Serlenga[‡]

January 18, 2017[§]

Abstract

We show that, when measuring inequality of opportunity with survey data, scholars face two types of biases. A well-known downward-bias, due to partial observability of circumstances that affect individual outcome, and an upward bias, which is the consequence of sampling variance. The magnitude of the latter distortion depends on both the empirical strategy used and the observed sample. We suggest that, although usually neglected in empirical contributions, the upward bias may be significant. We propose a simple criterion to select the best specification which balances between the two sources of bias. Our method is based on cross validation and can be easily implemented to survey data. In order to show how this method can improve our understanding of the inequality of opportunity measurement, we provide an empirical illustration based on income data of 26 European countries. Our evidence shows that estimates of inequality of opportunity are extremely sensitive to model selection. Alternative specifications lead to significant differences in the absolute level of inequality of opportunity and to a number of substantial countries' re-ranking. This in turn clarifies the need of an objective criterion to select the best econometric model when measuring inequality of opportunity.

Keywords: inequality of opportunity, model selection, variance-bias trade-off.
JEL: C52, D3, D63

[*]University of Florence & University of Bari, corresponding author: Dipartimento di Scienze per l'Economia e l'Impresa,Via delle Pandette, 32 - 50127 Firenze. paolo.brunori@uniba.it

[†]University of Bari.

[‡]University of Bari.

[§]This paper was presented at the worksop *Equality of opportunity and social mobility. Towards an international database*, 10-11 June 2016, University of Bari and at IT12 Winter School in Canazei 8-12 January 2017. We are grateful to all participants for their comments and suggestions. We also would like to thank Daniel Mahler and Raymond Schmelzer for their comments.

1 Introduction

The measurement of inequality of opportunity (IOp) is a growing topic in economics and, in the last two decades, the number of empirical contributions to this literature has exploded: see Ferreira and Peragine (2016) and Roemer and Trannoy (2015) for a review. The vast majority of these contributions are based on the approach proposed by Roemer (1998). This standard method consists in a two-step procedure: first, starting from a distribution of outcome (typically income or consumption), a counterfactual distribution is derived. This counterfactual distribution reproduces only unfair inequalities, i.e. inequalities due to circumstances beyond the individuals control, and does not reflect any inequality arising from choice and effort of the individual. Second, a suitable inequality measure is used to quantify inequality in the counterfactual distribution.

The empirical literature has extensively used two approaches to compute counterfactual distributions based on survey data: a parametric and a non-parametric one. One of the main drawbacks of both approaches is that, unless all circumstances beyond individual control that affect outcome are observable, they produce biased estimates of IOp. While the magnitude of this bias may be impossible to identify (Bourguignon et al., 2013), under some assumptions it can be shown that the sign of the bias is negative (Roemer, 1998; Ferreira and Gignoux, 2011; Luongo, 2011). This explains why IOp estimates are generally interpreted as lower-bound estimates of the "true" IOp, whereas the latter are interpreted as the estimates one would obtain if all circumstances were observable. Some authors have challenged the usefulness of those lower bound measures (Kanbur and Wagstaff, 2015; Balcazar, 2015). In particular, Balcazar (2015) and Ibarra et al. (2015) have suggested that the downward bias may lead to a substantial underestimation of the true level of IOp in empirical applications.

Typically, authors address this problem using rich data sources. In this case, the downward bias is minimized by either increasing the number of circumstances, as in Biörklund et al. (2012); or splitting circumstances in a broader partition of categories; or introducing interaction terms among categories, as in Hufe and Peichl (2015). Those empirical strategies reduce the downward bias by increasing the explained variability attributable to IOp. However, in this paper we emphasize that those procedures are not exempt by risk and might all lead to an upper distortion of the IOp estimates. Indeed, the reliability of the estimates would depend on both the number of circumstances and the types' partition but would also crucially depend on the sample distribution across types.

In both parametric and non-parametric approaches, we recognize a trade off between the downward bias resulting from the observability of circumstances and the upward bias related to the sampling variance of the estimated counterfactual distribution. Although the topic is not new to econometricians and practitioners, our concern over upward bias IOp estimates has been mostly neglected in the empirical literature of the IOp measurement. This is surprising because, as we show in the empirical section, such a distortion is likely to be far from negligible. We show that the magnitude of the upward distortion depends on the strategy used to obtain the counterfactual distribution. This is particularly straightforward when applying a parametric approach but can be easily generalized to the non parametric approach. A wrong choice of explanatory variables, as well as an inappropriate division in types' categories, might lead to distortions in both directions. We suggest that, when choosing among alternative specifications, scholars should opt for the best balance between the two sources of bias.

In this paper, after illustrating in detail the trade off and discussing its consequences for the measurement of IOp, we propose a method to select the best econometric specification, that is the specification that best minimizes both types of biases.

Our method is based on cross validation. The original sample is divided into a training set and a test set. The association between circumstances and outcome is first estimated on the training sample, under a large number of meaningful model specifications. Next, the derived coefficients are used to predict the outcome on the test sample. The specification selected is the model that minimizes the prediction errors in the test sample. This model selection technique is widely adopted in statistical learning; many routines have been developed and can be easily implemented in commonly used software.

In order to show the usefulness of our approach we apply our method to income data of 26 European countries using the EU Survey on Income and Living Conditions (EU-Silc) 2011 database. Our evidence shows that IOp estimates are extremely sensitive to model selection. Alternative specifications lead to significant differences in the absolute level of IOp and, in many cases, to substantial re-ranking of countries. Interestingly, our preferred specification generally differs from the typical model used in the literature. Hence, our estimates differ from estimates provided by other authors that have used the same data to estimate IOp.

The rest of the paper is organized as follows: Section 2 introduces the canonical model used to measure IOp, presents the estimation methods used to implement it and clarifies the two possible sources of distortion. Section 3 proposes a criterion to balance the trade-off between the two biases when selecting the specification to estimate IOp. Section 4 presents an empirical implementation while Section 5 concludes.

2 Downward and upward biased IOp

The canonical equality of opportunity model can be summarized as follows (see Ferreira and Peragine, 2016). Each individual in a society realizes an outcome of interest, y, by means of two sets of traits: circumstances beyond individual control, C, belonging to a finite set $\Omega = \{C_1, ..., C_J\}$, and a responsibility variable, e, typically treated as scalar. A function $g : \Omega \times \Re_+ \rightarrow \Re_+$ defines the individual outcome:

$$y = g(C, e)$$

For all $j \in \{1, ..., J\}$ let us denote by K_j the possible values taken by circumstance C_j and by $|K_j|$ the cardinality K_j. For instance, if C_j denotes gender, then $K_j = \{male, female\}$. We can now define a partition of the population into T types, where a type is a set of individuals who share exactly the same circumstances. That is, $T = \left|\Pi_{i=1}^{J} K_j\right|$. Let us denote by Y the overall outcome distribution.

IOp is then defined as inequality in the counterfactual distribution, $\tilde{Y}$, which reproduces all inequalities due to circumstances and does not reflect any inequality due to effort. A number of methods have been proposed to obtain $\tilde{Y}$ and, in general, the chosen method affects the resulting IOp measure (Ferreira and Peragine, 2016; Ramos and Van de gaer, 2015). In what follows, we focus on the *ex-ante* approach, introduced by Bourguignon et al. (2007) and Checchi and Peragine (2010), which is by far the most largely adopted method in the empirical literature (Brunori et al., 2013)[1]. This approach interprets the type-specific outcome distribution as the opportunity set of individuals belonging to each type. Then, a given value v_t of the opportunity set of each type is selected. Finally, $\tilde{Y}$ is obtained replacing the outcome of each individual belonging to type t with the value of her type v_t, for all $t = 1, ..., T$.

2.1 Counterfactuals estimation

Ex-ante IOp can be estimated either by a non parametric or a parametric approach. On the one hand, Checchi and Peragine (2010) propose to estimate $\tilde{Y}$ non-parametrically following the typical two stage method: (i) after partitioning the sample into types on the basis of all observable circumstances, they choose the arithmetic mean of type t, denoted by μ_t, as the value v_t of type t; (ii) for each individual i belonging to type t they define $\tilde{y}_i = \hat{\mu}_t$ - where $\hat{\mu}_t$ is

[1]Other well established approaches can be used to measure IOp. Approaches differ in how they define the principle of equal opportunity, and then in turn in the way the counterfactual distribution is constructed (Roemer, 1998; Lefranc et al., 2006; Fleurbaey and Shockaert, 2009; Checchi and Peragine, 2010). However, because the construction of these alternative counterfactual distributions generally requires the observation or identification of effort(an extremely difficult variable to measure) they are less frequently adopted in the empirical literature.

the sample estimate for μ_t - and measure inequality in $\tilde{Y}$.

On the other hand, Bourguignon et al. (2007) propose to measure *ex-ante* IOp by estimating $\tilde{Y}$ parametrically, as the prediction of the following reduced form regression:

$$y_i = \sum_{j=1}^{J} \sum_{k=1}^{K_j} \chi_{j_k} c_{i j_k} + u_i \tag{1}$$

where c_{ij_k} identifies each category of the observable characteristics by means of a dummy variable and χ_{j_k} is the corresponding coefficient[2]. The typical parametric approach consists in an ordinary least square regression where total outcome variability is explained by a linear combination of regressors with no interaction terms. It does not directly measure types' means but captures the variability explained by a given circumstance independently of all other characteristics considered. This restriction allows us to increase the degree of freedom of the regression, improving the reliability of the estimated counterfactual distribution. Understandably, such a parametric estimation has been proposed as a good alternative to the nonparametric when few observations are available (Ferreira and Gignoux, 2011; Ibarra et al., 2015).

Only recently, Hufe and Peichl (2015) discuss the importance of considering interaction terms in estimating IOp. Indeed, we notice that - when all explanatory variables are categorical - the parametric and the non parametric method coincide when the parametric counterfactual distribution is obtained by the prediction of a regression model where y is regressed on all possible combination of circumstances' values, i.e. all values of all regressors are interacted each other, obtaining a model with $(K \times J)^2$ dummies. In this case, each regressor captures the effect of belonging to one of all the possible circumstances' combination, which is the effect of belonging to a given type:

$$y_i = \sum_{t=1}^{T} \beta_t \pi_{it} + u_i \tag{2}$$

where π_{it} are T binary variables obtained by interacting all categories of all circumstances. In other cases, the corresponding IOp measures might be very different, and - by construction - the typical (linear) parametric approach (1) will explain less inequality than the non parametric (2) which allows variability to be explained also by the full set of interactions.

Here a trade off emerges: while the linear specification might be too restrictive, the inclusion of the full set of combinations among categories might lead to a very large sampling variance of the estimated counterfactual distribution. Especially when a limited number of observations

[2]Note that in principle one could have non-categorical regressors if cardinal circumstances are observed. However, to the best of our knowledge this is never the case. Even when a cardinal measure is available, such as parental income, authors tend to construct quantiles and use them as regressors (see for example Bjorklund et al., 2012).

is available for some types.

Following the same reasoning, the sampling variance of the estimated counterfactual distribution is also influenced by alternative partitions into categories of observed circumstances: a broadest partition might, again, lead to a larger variance in the case of a limited number of observations per types.

Indeed, the reliability of both parametric and non-parametric IOp estimates requires a sufficient number of observations characterized by each circumstance. In particular, the limitation is more severe in the case of the non parametric approach, as a sufficient number of observations for each combination of circumstances is required. In practice, this might represent a severe constraint as individuals are unlikely to be uniformly distributed across types and across categories partitions.

A typical argument arises when dealing with Western countries in which researchers observe both parental education and parental occupation as circumstances. Unfortunately, those variables are strongly correlated with each other and, usually, there are very few individuals whose parents are highly educated and employed in elementary occupations, or who have no education but work as managers. In order to overcome this drawback, scholars tend to limit the number of circumstances in the definition of types (using either parental education or parental occupation) or to aggregate the different values that a circumstance might take (using blue and white collars rather than more specific occupation categories) but those are obviously ad-hoc solutions which might greatly affect the shape of the counterfactual distribution and lead to ad-hoc IOp estimates. In what follows, we propose a statistical criterion to properly select among different model specifications or alternative categories' partitions.

2.2 Variance-bias trade-off in estimating IOp

A number of methodological contributions have shown that, if the 'true' set of circumstances is not fully observable, estimated IOp will be lower than the 'real' IOp (Ferreira and Gignoux, 2011; Luongo, 2011). This result follows from the assumption of orthogonality between circumstances and effort (see on this Roemer, 1998) and explains why IOp measures are generally interpreted as lower-bound estimates of IOp.

Typically, authors attempt to solve this problem by using rich datasets that contain the largest possible number of circumstances including outcome obtained during childhood (Bjorklund et al., 2012; Peichl et al., 2015). Recently, Niehues and Peichl (2014) endorse an extreme perspective and, by exploiting longitudinal datasets, they measure IOp including individual fixed effects among circumstances beyond individual control. This implies that any unobservable individual characteristics which persist overtime are considered a source of IOp. This method -

understandably - has been proposed as an 'upper-bound' estimates of the true IOp.

However, when using survey data, whenever one makes an effort to reduce the downward bias by increasing the number of circumstances or the number of categories in which circumstances might be split, she obtains a counterfactual distribution based on a finer partition in types. This, by construction, implies a smaller number of observations in each type[3]. This strategy might lead to higher between-group inequality and to a larger sampling variance when estimating the counterfactual distribution.

Surprisingly, the empirical literature on the estimates of IOp has so far neglected this second implication. Only recently it has been suggested that this issue may be an important aspect of IOp measurement. Brunori et al. (2016), working with Sub-Saharan African surveys, notice that the use of very detailed circumstances such as hundreds of 'village of birth' in Madagascar or hundreds of 'ethnic group' in Congo, tends to dramatically increase the IOp estimates.

However, it is important to notice that, when measuring inequality, higher sampling variance of the estimated distribution implies an upward biased IOp measure. This result is easily shown applying what Chakravarty and Eichhron (1994) proved for the case of inequality estimation when the variable of interest is measured with an error. The same result can be applied here: instead of the classical measurement error discussed in Chakravarty and Eichhron (1994), we consider a variable - the type mean - which is estimated with a higher sampling variance, the finer the partition in types.[4]

As far as the measurement of IOp is concerned, this result has two interesting consequences. Firstly, it states that if all circumstances are observable then IOp is upward biased. Secondly, whenever circumstances are not fully observable, two opposite distortions might bias our estimates and we can no longer assume that the estimated IOp is a lower-bound of the true IOp.

When sample size is large relative to the number of circumstances included in the model, the downward-bias is likely to be much larger. However, when the sample size is small relative to the number of types/regressors, it may be the case that the upward-bias prevails.

Yet, the absolute and relative sizes of the two biases depend on the sample size, the joint distribution of outcome and circumstances, and the model specification used to estimate the counterfactual distribution. That is to say, it is ultimately an empirical issue.

This discussion clarifies that, when estimating IOp, we should consider two different sources of distortion that bias our estimates in opposite directions. We cannot simply try to include the larger possible number of circumstances or to consider a broad partition of categories in order to minimize the downward distortion. Partial observability and sampling variance of the

[3]Or, if adopting a parametric approach, a regression with a larger number of controls and with fewer degrees of freedom.

[4]A formal proof is available in the appendix.

counterfactual distribution are both sources of bias that should be minimized. The following section proposes a simple method for choosing the best mode to measure IOp, i.e., the best way to exploit information contained in survey data, minimizing the sum of the two biases.

3 Model selection for measuring IOp

In this section we propose a method to select the most suitable model among the following alternatives: (i) the simple linear model (1) with the most parsimonious categories' partition, which provides the lowest extreme IOp estimates; (ii) a flexible model which includes the full number of combinations among categories - defined using the finest partition - (2), and leads to the highest value of the IOp estimates; (iii) all the intermediate specifications which only include subset of categories' combinations and alternative aggregations of characteristics' partitions.

This 'agnostic' approach, which does not impose any a priori restriction on the the effect of circumstances on outcome, may be inappropriate in other Contexts, such as when the aim is to test theoretical predictions of a theoretical model. However, since Bourgignon et al. (2007) IOp is measured using a reduced form model. Therefore, with no assumptions about the functional form of $g()$, to test for all possible models seems to be (when computationally affordable) the most appropriate choice.

In a statistical learning framework, we evaluate the variance bias trade-off in terms of model predictions. On the one hand, a more flexible model reduces the typical downward bias in IOp measurement and increases the prediction variance leading to upward bias. On the other hand, a more restricted model reduces the sampling variance and hence the upward bias, but suffers from omitted variable bias, the typical downward bias well known in the literature. In what follows, we propose to exploit the property of Mean Square Error (MSE) and choose the best model conditioned to available information by means of Cross Validation (CV). In a regression setting the MSE is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2$$

where y is the dependent variable, x are the regressors, and $i = 1, ..., n$ are the observations. For given out of sample observations y_0 and x_0, the MSE can always be decomposed in variance of $\hat{f}(x_0)$, square bias of $\hat{f}(x_0)$ and variance of the error term such as

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = Var(\hat{f}(x_0)) + \left[Bias\left(\hat{f}(x_0)\right)\right]^2 + Var(\epsilon)$$

where $\hat{f}(x_0)$ are the predictions. Since the variance of the error cannot be reduced, it turns out that in order to minimize the MSE, we need to minimize both the bias and the variance. A comparison among different specifications is performed by CV. In a CV procedure, the sample is randomly divided into k equal-sized parts. Leaving out part k (test sample), the model is fitted to the other $k - 1$ parts (training sample) whereas out-of- sample predictions are obtained for the left-out k^{th} part. CV compared with AIC, BIC and adjusted R^2 provide a direct estimate of the error. Overfitted models will have high R^2 values, but will perform poorly in predicting out-of-sample cases. CV is also useful to choose among alternative nonlinear specifications together with non nested models. For each specification, the average of the k MSEs is stored and the best specification is selected by minimizing it. This simple CV procedure is the widely acknowledged criterion that we propose to select the best specification among a number of possible alternatives: model (1), (2) and all the alternative specifications obtained both interacting only a subset of circumstances and defining categories using different partitions. Hence, our proposal is to estimate $\tilde{Y}$ with the model selected by CV. This is the best model which minimizes both sources of bias[5]. Note that this strategy might also imply the use of a different model for the same country in different time periods and, in general, each time the country's sample size differs.

A consequence of this is that, when comparing different countries in terms of IOp, we will compare measures obtained with different model specifications. This is somehow in contrast with what is generally proposed in the literature. When the same source of data is available for different countries, comparable measures of IOp have usually been computed using identical model specification for all countries, see Marrero and Rodriguez (2012), Brzezinski (2015) and Checchi et al. (2016). What we are suggesting here is a different approach. Comparable IOp measures should be calculated using the best performing model given the observable circumstances. To give a (simplistic) example, let us consider the comparison between France and Belgium in terms of IOp. It would make little sense to include among circumstances 'mother tongue' in France. Its inclusion implies a higher sampling variance and is unlikely to explain much of outcome inequality in the country. However, the same circumstance is likely to be one of the main sources of unequal opportunity in Belgium. In this case, we claim that the French counterfactual distribution should be estimated excluding 'mother tongue' from the set of regressors whereas, when estimating IOp in Belgium, this circumstance should be included.

We consider our method preferable when the intent is to compare the level of IOp in the two populations. On the one hand, it may seem unsatisfying, because measures are based on

[5] We are aware that the number of models to test explodes when circumstances are interacted. Moreover, some circumstances can enter into into the regression with alternative alternative level of detail, e.g. country of birth, region of birth, district of birth. When the level of detail to choose is not obvious this further increases the number of models that have to be checked. In these cases our method should be complemented with an algorithm able to restrict the number of models considered such as for example forward stepwise selection (Gareth et al., 2013).

different sets of information. On the other hand, the two measures are the two most reliable estimates of the effect of circumstances on outcome in the two populations.

Indeed, we believe that the specification used should differ, firstly because the set of available information may not be the same for the two populations. Secondly, and most importantly, because the nature of unequal opportunity, i.e. how circumstances affect individual outcomes, may differ in the two populations.

4 An empirical illustration

In order to clarify that our method is easily implementable and can substantially improve our understanding of IOp, we provide an empirical illustration based on the EU-Silc 2011 dataset. This dataset is the reference source for comparative statistics on income distribution in the European Union. Thanks to the special module on intergenerational transmission of poverty, the same data have been exploited for other estimates of IOp in the past, see Marrero and Rodriguez, 2012; Brzezinski, 2015; Checchi et al., 2016. In particular, in 2011 respondents provide information about their family of origin and socioeconomic background. The data contains information about 26 countries: Austria (AT), Belgium (BE), Bulgaria (BG), Czech Republic (CZ), Denmark (DK), Estonia (EE), Germany (DE), Great Britain (UK), Greece (GR), Finland (FI), France (FR), Hungary (HU), Italy (IT), Latvia (LV), Lithuania (LT), Luxembourg (LU), the Netherlands (NL), Norway (NO), Poland (PL), Portugal (PT), Romania (RO), Spain (ES), Slovakia (SK), Slovenia (SI), Sweden (SE) and Switzerland (CH).

Since our aim is to show how IOp estimates can be improved adopting the suggested model selection method, we compare our results with the measures obtained by Brzezinski (2015) and Checchi et al. (2016). Although both analyses use the EU-Silc 2011 data and an *ex-ante* parametric approach, the measures provided by those authors differ for at least two reasons: they use a different model specification and two alternative outcome definitions, e.g. Checchi et al. (2016) use individual disposal income whereas Brzezinski (2015) considers equivalized household disposable income. Hence, we apply our cross validation methodology to both outcome variables. In particular, following Checchi et al. (2016), we use individual disposal income[6] and restrict the sample to individuals aged between 30 and 60 who are either working full or

[6]The net individual income definition includes "(net) employee cash or near cash income" (variable PYN010) plus "(net) cash benefits or losses from self-employment" (variable PYN050 - negative values set equal to zero) plus "(net) non-cash employee income" (this variable is not available for all countries – variable PYN020). Capital incomes are excluded because they are only measured at household level, and it would be arbitrary to attribute them to household members. The disposable income definition add "(net) unemployment benefits" (variable PYN090), "(net) survivor' benefits" (variable PYN110), "(net) sickness benefits" (variable PYN120), "(net) disability benefits" (variable PYN130) and "(net) education-related allowances" (variable PYN140). Zero values are excluded from computation.

part-time, unemployed or fulfilling domestic tasks and care responsibilities. Further, we also perform our exercise on yearly equivalized disposable income for households whose head is between 26 and 50 years old, as specified in Brzezinski (2015).

We select gender, country of origin and family background as circumstances affecting individual incomes irrespective of individual responsibility. Those are all categorical variables and we consider them in both a parsimonious and a finer categories' partition. In the most parsimonious case only four binary variables are included in the regression equation: gender (male/female), country of origin (native/foreign), parental education (low/ high) and parental occupation (elementary/not elementary). In the broadest partition, we consider the following division: gender (male/female), country of origin (native/ EU foreign/ non EU foreign), mother and father occupation (coded in 10 values each)[7] and parental education (coded in 4 values each)[8]. Table 1 and 2 show the descriptive statistics of the most parsimonious case under both income definitions while Figure 1 shows the Gini IOp measures of three cases: (i) the linear, most parsimonious case (*low*), which is widely adopted in the literature, where categories are defined as binary variables and there are no interactions; (ii) the fully interacted model where the categories are defined in the widest option and are fully interacted (*up*); (iii) an intermediate measure computed from the best model selected by the CV method (*best*). This exercise is repeated for the two outcome variables.[9]

Both cases show that the three alternative measures clearly differ among one another, and in some cases (left) the best model and the linear model coincide. In other cases (right), the best model is far from the linear specification and rather close to the most flexible one. An immediate implication is that the position achieved in the countries' ranking clearly depends on the model specification chosen by the researcher. In light of those results, we suppose that, in measuring IOp, it is important to rely on a statistical method to select the best specification in order to avoid ad-hoc ranking. Next, we compare our *best* measure with alternative estimations obtained in other studies, which use the same EU-Silc data and the same definition of outcome variable, i.e. Checchi et al (2016) for disposal income and Brzezinski (2015) for the equivalized disposable income. We notice that the final assessment differs substantially in both cases. Figure 2 shows the rank correlation of our best measure and the two alternative estimates. Although the rank-correlation is clearly positive and significant, a number of countries lie outside the 45 degree line. Indeed, the re-ranking is substantial in a few cases. To give some examples in Checchi et

[7]ISCO-08: Armed forces occupations; Managers; Professionals; Technicians and associate professionals; Clerical support workers; Service and sales workers; Skilled agricultural, forestry and fish; Craft and related trades workers; Plant and machine operators; Elementary occupations.

[8]Could neither read nor write; Low level (pre-primary, primary education); Medium level (upper secondary education); High level (first stage of tertiary education).

[9]Tables 3 and 4 in the Appendix contains IOp estimates and relative bootstrapped standard errors.

Figure 1: IOp in 26 European countries under different model specifications

Source: EU-SILC, 2011 . Note: The Figure shows each country's IOp measure with the three alternative methods: (i) the linear, most parsimonious case (*low*), (ii) the fully interacted model (*up*); (iii) the best model selected (*best*) using individual equivalized income. Error bars show 95% bootstrap confidence intervals.

al. (2016) Poland ranks 20th and the Netherlands rank 5th, whereas accordingly with our *best* measure they rank 6th and 9th, respectively. As well in Brzezinski (2015), Belgium lies at the 16th position whereas, if the best specification is adopted, it would be 7th.

We believe that this exercise provides convincing evidence that the variance-bias trade-off in IOp measurement is far from negligible in empirical applications. It is therefore crucial to introduce a widely accepted statistical criterion to select the best model among a very large number of possible specifications.

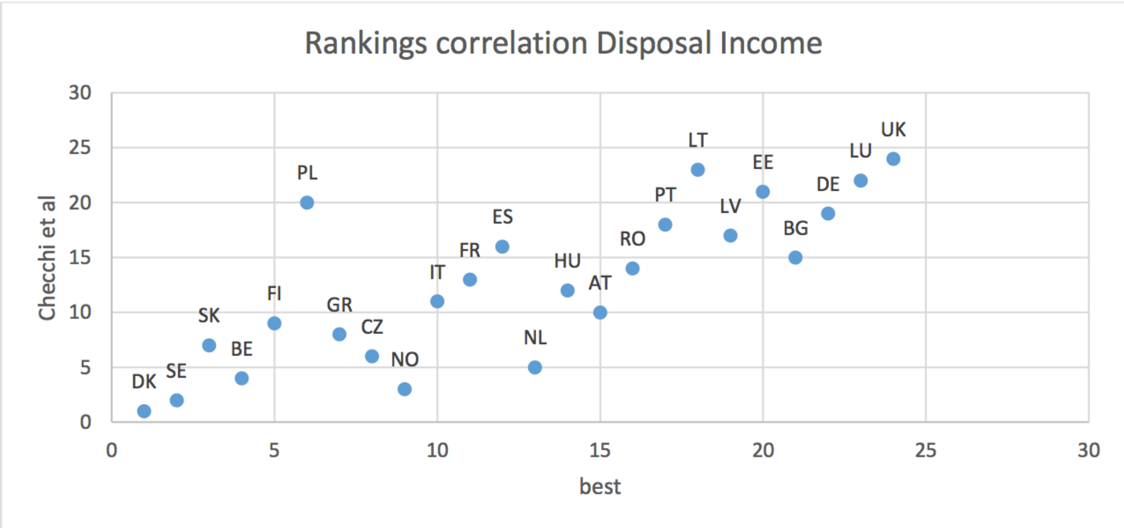
5 Conclusions

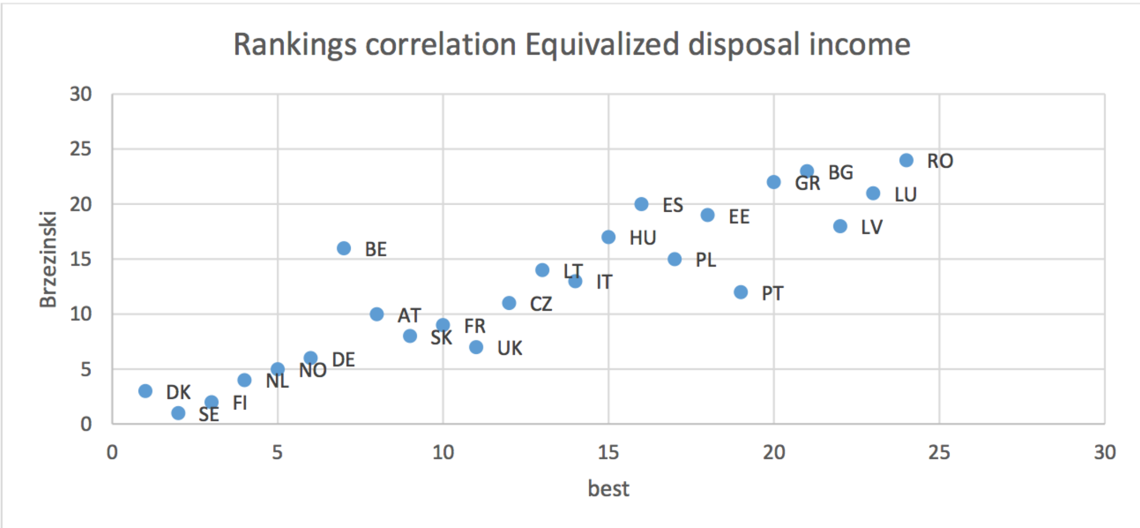
Scholars are well aware that the IOp estimates are mostly downward biased. This is a consequence of the partial observability of circumstances beyond individual control that affect individual outcome. However, since IOp is measured as inequality in a counterfactual sample distribution, a second possible source of bias might be related to the sampling variance of the estimated counterfactual distribution. In this paper, we discuss this further source of bias, which has surprisingly been neglected by the empirical literature on IOp measurement. We show that it implies an upward bias of IOp which challenges the interpretation of IOp estimates as lower-bound estimates of the real level of IOp.

We stress that, because the empirical specification used to estimate IOp largely influences its magnitude, we need a reasonable criterion to choose among alternative models. We suggest that this criterion should minimize the two sources of bias.

We interpret this problem as a typical variance-bias trade-off and propose to adopt a simple cross validation method to find the best fitting model. Finally, we show the empirical relevance of our intuition and implement the proposed method to measure IOp in 26 European countries. Our empirical evidence clarifies that the choice of the model specification largely affects the estimated IOp and demonstrates the importance of having a widely accepted criterion to identify the best possible specification.

Figure 2: IOp in 26 European countries estimates from different studies





Source: EU-SILC, 2011 . Note: the Figure shows the rank correlation of countries in terms of IOp. Our best model specification is compared with Checchi et. al. (2016) and Brzezinski (2015)

References

- Balcazar, C. (2015), "Lower bounds on inequality of opportunity and measurement error", *Economics Letters*, 137: 102-105.

- Biörklund, A., Jäntti A. and Roemer J. (2012), "Equality of Opportunity and the Distribution of Long-Run Income in Sweden", *Social Choice and Welfare*, 39: 675-696.

- Bourguignon, Ferreira F., and Menendez M., (2007), "Inequality of Opportunity in Brazil", *Review of Income Wealth*, 53: 585-618.

- Bourguignon, Ferreira F., and Menendez M., (2013), "Inequality of Opportunity in Brazil: A Corrigendum", *Review of Income Wealth*, 59: 551-555.

- Brzenziński M. 2015. 'Inequality of opportunity in Europe before and after the Great Recession', Working Paper n. 2/2015 (150). Faculty of Economic Sciences, University of Warsaw.

- Brunori P., Palmisano, F., Peragine, V. (2015), "Inequality of opportunity in Sub Saharan Africa", World Bank Policy Research Working Papers 7782.

- Brunori, P., Ferreira F., Peragine, V. (2013), "Inequality of opportunity, income inequality and mobility: some international comparisons", in *Getting Development Right: Structural Transformation, Inclusion and Sustainability in the Post-Crisis Era* edited by E. Paus. Palgrave Macmillan.

- Checchi, D., and Peragine V. (2010), "Inequality of Opportunity in Italy", *Journal of Economic Inequality*, 8: 429-450.

- Checchi, D., Peragine V., Serlenga L., (2016), "Inequality of Opportunity in Europe: Is There a Role for Institutions?" , in Cappellari L., Polachek S., Tatsiramos K. (ed.) *Inequality: Causes and Consequences* , Research in Labor Economics, Volume 43, Emerald.

- Ferreira, F., and Gignoux, J. (2011), "The measurement of inequality of opportunity: theory and an application to Latin America", *Review of Income and Wealth*, 57: 622-657.

- Ferreira F., and V. Peragine. (2016). "Equality of Opportunity: Theory and Evidence", in *Oxford Handbook of Well-Being and Public Policy* edited by M. Adler and M. Fleurbaey. Oxford University Press.

- Gareth J., Witten D., Hastie T., Tibshirani R. (2013). “An Introduction to Statistical Learning with Applications in R”, Springer, New York.

- Hufe P and A. Peichl (2015). “Lower Bounds and the Linearity Assumption in Parametric Estimations of Inequality of Opportunity”, IZA Working Papers, DP No. 9605.

- Kanbur, R., Wagstaff, A. (2015), “How useful is inequality of opportunity as a policy construct?”, *CEPR Discussion Paper* DP10508.

- Ibarra L., Martinez C., Adan L. (2015), “Exploring the sources of downward bias in measuring inequality of opportunity.” Policy Research working paper; no. WPS 7458. Washington.

- Luongo P. (2011),“The Implication of Partial Observability of Circumstances on the Measurement of Inequality of Opportunity”, in Rodriguez, J. eds. , *Research on Economic Inequality*, 19: 23-49.

- Marrero, G., and J. Rodrguez. (2012). “Inequality of Opportunity in EuropeÕ. *Review of Income and Wealth*, vol. 58: 597-621.

- Niehues, J., and Peichl A. (2014), “Upper bounds of inequality of opportunity: theory and evidence for Germany and the US”, *Social Choice & Welfare*, 43: 63-79.

- Peichl A., Hufe P., Roemer J., Ungerer M. (2015), “Inequality of Income Acquisition: The Role of Childhood Circumstances” ZEW - Centre for European Economic Research Discussion Paper No. 15-084.

- Ramos, X., Van de gaer, D. (2015), “Measurement of Inequality of Opportunity Based on Counterfactuals”, *Paper presented at the 6th ECINEQ meeting.* University of Luxembourg, 2015.

- Roemer, J. (1998), “Equality of Opportunity”, Cambridge, MA: Harvard University Press.

- Roemer J., and Trannoy A ., (2015), “Equality of Opportunity”, In *Handbook of Income Distribution* Edited by Anthony B. Atkinson and Francois Bourguignon. vol. 2. Elsevier.

Table 1: Descriptive statistics

	AT	BE	BG	CH	CZ	DE	DK	EE	GR	ES	FI	FR	HU
obs	6097	5892	6989	7433	8538	12342	5781	5233	5990	15174	9550	10859	13067
median disposable income	19946	20700	2454	43648	7214	17608	33506	5113	12902	13800	27613	19658	4322
female	0.53	0.52	0.51	0.54	0.52	0.54	0.53	0.52	0.52	0.52	0.5	0.52	0.54
age (years)	45.36	45.03	46.19	45.73	45.56	46.5	46.66	45.77	45.26	45.07	46.52	45.7	46
foreign	0.16	0.17	0.01	0.25	0.03	0.06	0.07	0.13	0.10	0.10	0.04	0.10	0.01
parental occupation: elementary	0.15	0.09	0.21	0.09	0.13	0.06	0.05	0.17	0.09	0.19	0.08	0.27	0.26
parental education: low	0.37	0.51	0.51	0.24	0.60	0.11	0.10	0.34	0.78	0.83	0.49	0.76	0.61
	IT	LT	LU	LV	NL	NO	PL	PT	RO	SE	SI	SK	UK
obs	3648	20652	5296	6632	4654	11179	4927	15238	5755	7699	6469	12926	6712
median disposable income	17972	17944	3791	31163	20946	31864	45198	4844	7899	2092	24082	11102	14876
female	0.51	0.52	0.54	0.52	0.52	0.53	0.51	0.52	0.53	0.52	0.52	0.51	0.53
age (years)	45.67	45.32	47.58	44.57	46.52	46.22	45.5	45.96	46.03	46.18	45.67	45.58	45.62
foreign	0.09	0.07	0.52	0.13	0.06	0.09	0.01	0.08	0.01	0.14	0.12	0.01	0.12
parental occupation: elementary	0.15	0.37	0.13	0.27	0.04	0.06	0.16	0.19	0.13	0.01	0.11	0.26	0.16
parental education: low	0.76	0.58	0.51	0.41	0.38	0.23	0.49	0.93	0.85	0.34	0.67	0.35	0.55

Source: EUSILC (2011)

Table 2: Descriptive statistics

	AT	BE	BG	CH	CZ	DE	DK	EE	ES	FI	FR	GR	HU
hline obs	22340	22586	3135	37287	7755	21290	30491	6181	13238	23846	20630	11080	4576
median equivalized income	3844	3401	2323	4152	3928	6937	1242	3108	7991	2186	6842	2722	6753
female	0.48	0.50	0.50	0.48	0.43	0.49	0.49	0.49	0.49	0.53	0.48	0.51	0.48
age (years)	39	39	41	40	39	40	40	40	40	38	39	40	39
foreign	0.18	0.18	0.01	0.25	0.03	0.06	0.07	0.08	0.13	0.06	0.09	0.13	0.01
parental occupation: elementary	0.14	0.12	0.22	0.10	0.15	0.05	0.01	0.14	0.20	0.19	0.28	0.09	0.24
parental education: low	0.29	0.41	0.42	0.20	0.51	0.07	0.05	0.22	0.81	0.30	0.69	0.74	0.53
	IT	LT	LU	LV	NL	NO	PL	PT	RO	SE	SI	SK	UK
obs	17171	4213	31655	4511	23005	38392	5025	8476	2152	22899	12972	6523	19484
median equivalized income	10533	2119	4203	2640	3769	1913	7645	2825	2611	459	3489	3413	3959
female	0.49	0.48	0.49	0.45	0.48	0.54	0.47	0.48	0.51	0.45	0.49	0.47	0.45
age (years)	40	42	39	40	40	39	40	41	40	38	38	41	40
foreign	0.12	0.05	0.58	0.09	0.06	0.10	0.00	0.10	0.00	0.14	0.09	0.01	0.14
parental occupation: elementary	0.17	0.34	0.15	0.25	0.09	0.11	0.17	0.18	0.15	0.06	0.32	0.22	0.18
parental education: low	0.72	0.42	0.49	0.29	0.29	0.17	0.35	0.91	0.79	0.34	0.58	0.25	0.50

Source: EUSILC (2011)

Table 3: IOp (Gini) esitmates - Individual disposal income

	up	best	low
AT	0.1980	0.1338	0.1336
	(0.0033)	(0.0014)	(0.0013)
BE	0.1734	0.1095	0.0858
	(0.0048)	(0.0021)	(0.0013)
BG	0.1911	0.1609	0.1290
	(0.0028)	(0.0021)	(0.0011)
CH	0.2694	0.1879	0.1875
	(0.0035)	(0.0011)	(0.0015)
CZ	0.1497	0.1219	0.1142
	(0.0019)	(0.0013)	(0.0009)
DE	0.2070	0.1612	0.1313
	(0.0026)	(0.0014)	(0.0008)
DK	0.1452	0.0962	0.0655
	(0.0043)	(0.0022)	(0.0010)
EE	0.2168	0.1591	0.1202
	(0.0034)	(0.0021)	(0.0017)
ES	0.2073	0.1265	0.1114
	(0.0043)	(0.0020)	(0.0015)
FI	0.1882	0.1102	0.0850
	(0.0042)	(0.0018)	(0.0011)
FR	0.1826	0.1245	0.1112
	(0.0029)	(0.0012)	(0.0011)
GR	0.1731	0.1148	0.0995
	(0.0053)	(0.0023)	(0.0023)
HU	0.1732	0.1337	0.1172
	(0.0016)	(0.0010)	(0.0008)
IT	0.1729	0.1225	0.1122
	(0.0023)	(0.0013)	(0.0011)
LT	0.1772	0.1517	0.0671
	(0.0040)	(0.0033)	(0.0005)
LU	0.2422	0.1748	0.1543
	(0.0037)	(0.0030)	(0.0023)
LV	0.2346	0.1584	0.1100
	(0.0033)	(0.0021)	(0.0012)
NL	0.2006	0.1323	0.1195
	(0.0042)	(0.0019)	(0.0013)
NO	0.2038	0.1220	0.1043
	(0.0047)	(0.0017)	(0.0010)
PL	0.1463	0.1145	0.0902
	(0.0019)	(0.0010)	(0.0007)
PT	0.2131	0.1427	0.0844
	(0.0047)	(0.0023)	(0.0024)
RO	0.1949	0.1369	0.1365
	(0.0028)	(0.0020)	(0.0022)
SE	0.2166	0.1008	0.0609
	(0.0105)	(0.0051)	(0.0023)
SI	0.1451	0.0825	0.0662
	(0.0030)	(0.0010)	(0.0008)
SK	0.1277	0.1078	0.0851
	(0.0020)	(0.0013)	(0.0007)
UK	0.2619	0.1833	0.1640
	(0.0040)	(0.0018)	(0.0014)

Source: EUSILC 2011. Notes: IOp (Gini) estimates derived from (i) the linear most parsimonious case (low); (ii) the fully interacted model (up); (iii) the best model selected (best), respectively. Bootstrapped standard errors in parenthesis. .

Table 4: IOp (Gini) esitmates - Individual disposal equivalent income

	up	best	low
AT	0.1295	0.0653	0.0505
	(0.0032)	(0.0015)	(0.0011)
BE	0.1326	0.0631	0.0517
	(0.0038)	(0.0015)	(0.0010)
BG	0.1899	0.1554	0.1232
	(0.0038)	(0.0027)	(0.0020)
CH	0.1596	0.0956	0.0456
	(0.0035)	(0.0013)	(0.0008)
CZ	0.1165	0.0866	0.0661
	(0.0020)	(0.0014)	(0.0005)
DE	0.1141	0.0535	0.0231
	(0.0020)	(0.0006)	(0.0003)
DK	0.1097	0.0126	0.0095
	(0.0048)	(0.0012)	(0.0007)
EE	0.1709	0.1241	0.0632
	(0.0035)	(0.0022)	(0.0012)
ES	0.1909	0.1123	0.0900
	(0.0041)	(0.0020)	(0.0018)
FI	0.1300	0.0386	0.0166
	(0.0035)	(0.0008)	(0.0003)
FR	0.1284	0.0728	0.0572
	(0.0024)	(0.0008)	(0.0007)
GR	0.1881	0.1329	0.1138
	(0.0056)	(0.0038)	(0.0035)
HU	0.1412	0.1083	0.0922
	(0.0016)	(0.0009)	(0.0006)
IT	0.1595	0.1022	0.0926
	(0.0026)	(0.0013)	(0.0012)
LT	0.1707	0.0919	0.0767
	(0.0042)	(0.0014)	(0.0009)
LU	0.1935	0.1787	0.1119
	(0.0031)	(0.0028)	(0.0010)
LV	0.2224	0.1574	0.0980
	(0.0043)	(0.0027)	(0.0017)
NL	0.1169	0.0437	0.0341
	(0.0029)	(0.0011)	(0.0007)
NO	0.1183	0.0451	0.0237
	(0.0035)	(0.0010)	(0.0007)
PL	0.1475	0.1209	0.0855
	(0.0018)	(0.0008)	(0.0006)
PT	0.1886	0.1259	0.0517
	(0.0046)	(0.0024)	(0.0028)
RO	0.2115	0.1833	0.1384
	(0.0036)	(0.0025)	(0.0027)
SE	0.1514	0.0350	0.0317
	(0.0084)	(0.0026)	(0.0022)
SI	0.1077	0.0706	0.0637
	(0.0019)	(0.0008)	(0.0006)
SK	0.0996	0.0715	0.0324
	(0.0021)	(0.0014)	(0.0006)
UK	0.1701	0.0823	0.0535
	(0.0035)	(0.0011)	(0.0007)

Source: EUSILC 2011. Notes: IOp (Gini) estimates derived from (i) the linear most parsimonious case (low); (ii) the fully interacted model (up); (iii) the best model selected (best), respectively. Bootstrapped standard errors in parenthesis. .

5.1 Upward bias when estimating IOp with survey data

Chakravarty and Eichhron (1994) distinguish between the true distribution of income, y, and the observed one $\tilde{y}$ where $\tilde{y} = y + e$ and e is commonly defined as measurement error such that $e \sim iid(0, \sigma^2)$. By considering a strictly concave von Neumann-Morgenstern utility function of the individuals, U, they prove by analogy that if, we measure inequality $I(\tilde{y})$ with an inequality index I that satisfies symmetry and Pigou-Dalton transfer principle, then the inequality of the true counterfactual distribution is smaller than what observed in the sample.

In the case of IOp - assuming no measurement error and assuming to observe all circumstances beyond individual control affecting outcome - we know that our counterfactual distribution $\tilde{Y}$ is estimated with some degree of uncertainty. The uncertainty depends on the available degree of freedom. A finer partition of the population and, therefore, smaller types' sample size, leads to fewer degree of freedom and larger sampling variance. This implies that the IOp in the sample is smaller the IOp estimated in a sample. The finer the partition in types the larger the upward bias.

More in details, if μ_t is the type mean when the number of observations within types is small, we expect a biased estimates of sample mean, such that $\tilde{\mu}_t = \mu_t + \eta$ where $\tilde{\mu}_t$ is the estimated type mean, μ_t is the "true" parameter and η is the standard error of $\tilde{\mu}_t$, i.e. $\frac{\sigma}{\sqrt{N_t}}$. Simulations prove that the error component leads to a positive distortion and by construction converges to zero as $N_t \to \infty$. Following Chakravarty and Eichhron (1994) we can easily prove that between inequality derived by a larger partition of the population is an overestimation of that derived by smaller (and more representative) ones.

Assuming that U is strictly concave by Jensen's inequality, we have

$$E\left(U\left(\tilde{\mu}_t|\mu_t\right)\right) < U\left(E\left(\tilde{\mu}_t|\mu_t\right)\right)$$

given that

$$\begin{aligned} \tilde{\mu}_t &= \mu_t + \eta \qquad (3) \\ &= E\left(\tilde{\mu}_t|\mu_t\right) + \overbrace{\left(\tilde{\mu}_t - E\left(\tilde{\mu}_t|\mu_t\right)\right)}^{\eta} \\ \tilde{\mu}_t - \eta &= E\left(\tilde{\mu}_t|\mu_t\right) \text{ from (3) } \tilde{\mu}_t - \eta = \mu_t \end{aligned}$$

and

$$U\left(E\left(\tilde{\mu}_t|\mu_t\right)\right) = U\left(\mu_t\right) \qquad (4)$$

Then

$$E\left(U\left(\tilde{\mu}_t|\mu_t\right)\right) < U\left(\mu_t\right)$$

Taking expectation of both sides of (4) with respect μ_t, we get

$$E\left(U\left(\tilde{\mu}_t\right)\right) < U\left(E\left(\mu_t\right)\right) \tag{5}$$

Given that $\tilde{\mu}_t$ and μ_t asymptotically - as $N_t \to \infty$ - have the same mean and U is strictly concave.

Therefore, given the circumstances observed, if IOp is estimated by applying an inequality index satisfying symmetry and Pigou-Dalton transfer principle, IOp estimates are an upward biased estimate of the real between-type inequality. The bias is monotonically increasing with the number of observed circumstances and is monotonically decreasing with the sample size.